General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

APF76-MSFC-02127

APF76-MSFC-02127

Technical Report

DATA BASE MANAGEMENT STUDY
INTERIM REPORT

(NASA-CR-150167) DATA BASE MANAGEMENT STUDY Interim Report (Teledyne Brown Engineering) 62 p HC A04/MF A01 CSCL 09B

N77-15671

Unclas
G3/60 12515

May 1976

FEB 1977
RECEIVED
NASA STI FACILITY
INPUT BRANCH

TELEDYNE
BROWN ENGINEERING

Cummings Research Park • Huntsville, Alabama 35807

TECHNICAL REPORT
APF76-MSFC-02127

DATA BASE MANAGEMENT STUDY INTERIM REPORT

May, 1976

Prepared for
Data Systems Laboratory
George C. Marshall Space Flight Center
Huntsville, Alabama

Contract No. NAS8-31488

Prepared by
Advanced Projects Division
Teledyne Brown Engineering
Huntsville, Alabama

ABSTRACT

This interim report presents the results of work performed by Teledyne Brown Engineering (TBE) under NASA contract NAS8-31488. It describes data base management techniques and applicable equipment and contains recommendations which will assist potential NASA data users in selecting and using appropriate data base management tools and techniques. The report is organized to present the primary areas of investigation.

Part 1 surveys classes of currently available data processing equipment ranging from basic terminals to large minicomputer systems, as they apply to the needs of potential SEASAT data users. It also includes cost and capabilities projections for this equipment through 1985. Part 2 describes a test conducted by TBE of a typical Data Base Management System, the results of this test, and recommendations to assist potential users in determining when such a system is appropriate for their needs. The representative system tested was UNIVAC's DMS 1100. Part 3 presents the requirements that potential users will have for NASA data. Based on these requirements and the data reporting plans for several NASA programs, it also discusses the resulting impact on user equipment needs.

Approved:

RR Parker

R. R. Parker, Ph.D
Manager
Data System Projects

R S McCarter

R. S. McCarter
Vice President

TABLE OF CONTENTS

Part 1 - Survey and Capabilities Projections for SEASAT User Data Systems

Page

1. INTRODUCTION 1-1

2. APPROACH 2-1
   2.1 Survey of Currently Available Equipment 2-1
   2.2 Capability and Cost Projection 2-3

3. TELECOMMUNICATIONS SYSTEMS 3-1

4. MICROCOMPUTER BASED SYSTEMS 4-1

5. SMALL MINICOMPUTER BASED SYSTEMS 5-1

6. LARGE MINICOMPUTER BASED SYSTEMS 6-1

7. EQUIPMENT PROJECTIONS THROUGH 1985 7-1
   7.1 Capabilities by Equipment Categories 7-2
      7.1.1 Telecommunications Systems 7-3
      7.1.2 Microcomputer Systems 7-4
      7.1.3 Small Minicomputer Systems 7-6
      7.1.4 Large Minicomputer Systems 7-7
   7.2 System Cost Projections 7-8

Part 2 - DMS 1100 Test and Evaluation

1. INTRODUCTION 1-1

2. TEST PROCEDURE AND RESULTS 2-1
   2.1 Test Description 2-1
      2.1.1 Environment 2-1
      2.1.2 Files 2-2
      2.1.3 Operations on Files 2-3

TABLE OF CONTENTS CONTINUED

Page

2.2 Quantitative Test Results 2-3

3. ANALYSIS OF TEST RESULTS. 3-1

3.1 Ease Of Use. 3-1

3.1.1 Documentation 3-2

3.1.2 Program Diagnositcs 3-2

3.1.3 Comments on System Design 3-3

3.2 Machine Resources. 3-3

4. RECOMMENDATIONS 4-1

4.1 When Is a DBMS Appropriate 4-1

4.2 DMS 1100 4-3

Part 3 - NASA Data Users' Requirements For Processing Equipment

1. INTRODUCTION 1-1

2. USERS' DATA NEEDS 2-1

2.1 Categorization of Data Needs 2-1

2.2 Examples of Potential Users. 2-2

3. INFLUENCE OF DATA NEEDS ON EQUIPMENT REQUIREMENTS . . 3-1

3.1 Urgency of Data. 3-1

3.2 Processing Required. 3-1

3.3 Presentation of Data 3-2

3.4 Retention Requirements 3-3

PART 1

SURVEY AND CAPABILITIES PROJECTIONS FOR SEASAT USER DATA SYSTEMS

Part 1 of this document surveys minicomputer and microcomputer equipment available to potential users of SEASAT data. Complete processing systems appropriate for SEASAT users are described, each of which can provide local access to central data files and a particular level of local processing power. Projections of system capabilities and costs through 1985 are also presented. Just as user needs can be categorized by requirements for local storage, processing power, and equipment sophistication, the systems described in this report have been categorized for ease of reference. The categories presented are: telecommunications terminals (communication only with no local processing), microcomputer based systems (provide basic processing capability), small minicomputer systems (added processing power and graphics capability), and large minicomputer systems (powerful systems able to support large processing projects and networks of users).

1. INTRODUCTION

The potential users of SEASAT data will vary widely in terms of the data processing their applications will require, the resources they are able to commit to these projects, and their own technical capabilities. To assist them in the procurement and analysis of SEASAT data, they will require data processing equipment ranging from the most basic terminals to sophisticated, relatively independent computer systems. This report describes and predicts the future of applicable data processing and communications equipment from simple communications terminals through large, stand-alone minicomputer systems.

The equipment discussed in this report has been categorized into four classes of configurations. Each configuration provides users of SEASAT data with a different level of processing capabilities. Such aspects of user requirements as the amount of data which must be manipulated, the amount of processing to be performed, and the computer sophistication necessary for their tasks were used to define these categories. The following section on the survey and projection approaches used for this report describes these classes and indicates their applicability to SEASAT data processing. Potential users should first examine this next section to determine the type of equipment of most interest to them. They should then examine the corresponding portions of Sections 3, 4, 5, and 6, which describe these configurations in further detail and present the results of our survey of current, representative systems. Section 7, which presents capabilities and cost projections for the equipment through 1985, can be used to help in formulating future equipment acquisition plans.

2. APPROACH

The following subsections describe the approaches used in performing this survey of currently available equipment and in predicting the future of this type of equipment. Also noted are significant assumptions used during the course of these activities.

2.1 SURVEY OF CURRENTLY AVAILABLE EQUIPMENT

The categorization of current equipment used in this report is summarized in Table 2-1. This table also indicates the applicability of each category to SEASAT data users. The range of equipment available in each category makes it impossible to define distinct boundaries between categories. For example, a sophisticated microprocessor system may actually be more powerful than an inexpensive minicomputer based system. Thus, users should have their requirements clearly in mind and be prepared to examine a wide variety of equipment to determine what suits their needs best.

In the following four sections, equipment representative of each category is described. A major objective in choosing this representative equipment was to demonstrate the variety of capabilities and technologies available in each category. In most cases, this equipment is from the larger, more established vendors. These vendors generally have the most complete systems and offer widely available service. These advantages, however, do have their cost, usually evident in total system price. Often by mixing mainframe and peripheral vendors, users can obtain superior performance and lower cost. Users must balance the simplicity of inter-

TYPE OF EQUIPMENT	BASIC SYSTEM COMPONENTS	BASIC SYSTEM APPROX. PRICE	OPTIONAL COMPONENTS	SEASAT DATA PROCESSING CAPABILITIES
1 - Telecommunications Only (See Section 3)	• Typewriter-like or CRT Terminal	$1000 to $4,800		• Interractive Inquiry To Locate Appropriate Data • Limited Receipt Of SEASAT Data
2 - Microcomputer Based Terminal With Some Local Processing Capability (See Section 4)	• Microprocessor: For Control and Processing • System Monitor: Key-board and Display • Local Data Storage: Floppy Disk or Cassette Tape	$9000 to $16,000	• Printer • Graphics Equipment	Same As Above Plus: • Formatting Data Requests And Decoding Responses From Central System To Simplify Communication • Automated Storage And Examination of Limited Amounts of SEASAT Data
3 - Small Minicomputer Based System With Graphics Capability (See Section 5)	• Minicomputer (16-Bit Word Size) • Small Mass Storage-Disk and Mag. Tape • Slow Printer and Card Reader • Graphics CRT	$50,000 to $87,000	• Hardcopy device for graphic out-put • Higher speed I/O devices	Same As Above Plus: • Graphical Inquiry And Examination of SEASAT Data • More Processing power
4 - Large Minicomputer Based System With Graphics Capability (See Section 6)	• Minicomputer (32-bit Word Size) with large Memory • Med. to large mass storage-disk and mag. tape • Med. speed I/O devices-printer and card reader • Graphics CRT • Hardcopy device for Graphics output	$137,000 to $153,000	• Higher Speed I/O Devices • Color CRT	Same As Above Plus: • Able To Support Extensive Stand-Alone Processing Of Large Volumes of SEASAT Data, Including Graphics Applications • Able To Support Multiple Users or Even A Network Of Terminal Users

TABLE 2-1 SUMMARY OF CURRENT EQUIPMENT CATEGORIES AND CAPABILITIES

connecting components and the clear assignment of maintenance responsibility in single vendor systems aganist the potential cost savings of multiple vendor configutations. The prices quoted in the following sections are list prices for single unit purchases. Comparison pricing of competitive products at the actual time of purchase may disclose price changes or negotiable discounts.

2.2 CAPABILITY AND COST PROJECTIONS

Currently, computers are referred to as micro, mini, or just plain computers (milli and maxi have also recently appeared as vendors attempt to stress the uniqueness of their products). The basis for this classification has been the technologies used (e.g. LSI in microcomptuers), the system cost, the level of support available, and in some cases the historic product lines of manufacturers. As the gaps between the categories are filled by new products; as various technologies and levels of support become available on all sizes of computers; and as costs respond to technological advances, the fate of these individual classes of equipment will become impossible to follow. Similarly, it is difficult to choose a level of processing power and to follow that level of processing over time since expected advances in capabilities and even approaches to processing will soon outdate current standards. Therefore, the types of SEASAT user applications appropriate for the equipment categories defined in the previous subsection will serve to define the categories of future equipment projected by this report through 1985.

Although computer peripherals often include mechanical components, computer systems are predominately electrical devices. Thus their cost

and capability trends are expected to be similar to historic trends for other electronic devices. Much of the basis for the projections in this report is the dramatic design and manufacturing advances historically present with useful electronic devices and the subsequent reductions in the price of these devices. Also, computers themselves have existed long enough to have established some of their own cost and performance trends, which were used in formulating the projections presented in this report. Projections based on such methods, however, assume a relatively continous refinement and development of improved products. Dramatic breakthroughs which cause major deviation from the current trends in technology or costs cannot be anticipated. All costs in this report are expressed in terms of 1976 dollars.

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

3. TELECOMMUNICATIONS SYSTEMS

A basic telecommunications terminal can provide users with a minimal cost point of access to SEASAT data. Such a terminal can not process data itself, but can request processing by a central site computer and can receive responses. Depending upon central site capabilities, users may be provided with processing such as the following:

1) Searches of available SEASAT data for that meeting specific criteria (i.e. all coverage of a certain location on a specified date with less than 15% cloud cover)

2) Central site processing of SEASAT data (i.e. determining wind speeds from specified SEASAT data)

3) Summary output of either of the two above tasks transmitted to the terminal or dumped to hardcopy media at the central site and mailed to a specified location.

The communication process itself can range from a strictly formatted procedure to an interactive system where the central site assists the user in defining his requests. Because of their lack of local processing capabilities, all data to be displayed by one of these terminals must be transmitted in the form it is to be displayed. This, coupled with the relatively slow speed of these devices (many function at 10 to 30 characters per second) make them practical for receipt of only limited amounts of data. In additon, these devices generally can only handle alphanumeric data, which rules out display of graphic images derived from SEASAT data. Since the bulk of processing for such terminals is handled by a central site, inquiry from a large number of these devices produces a heavy processing load at this central site. The low cost of

such terminals therefore must be balanced aganist the costs for telephone service and powerful central site processing capabilities.

The specific devices falling into this category are the "dumb" typewriter-like and CRT terminals (so named because of their lack of processing power). The most familiar typewriter-like terminal is the teletype. The CRT's resemble small television sets with a typewriter keyboard attached. In order to communicate with a central computer site, these devices also require a modem (integral part of some of the terminals) to interface with voice-grade telephone lines. Some optional features available to users include editing capability to allow modification of data before transmission, CRT hardcopy devices, and on-site storage to accumulate requests or replies. The primary advantage of typewriter-like terminals is that hardcopy output is a natural by-product, whereas it is expensive to produce from a CRT. In contrast, CRT's are generally faster and much quieter to operate. Table 3-1 summarizes equipment representative of this category.

TABLE 3-1 TELECOMMUNICATIONS TERMINALS

MANUFACTURER AND MODEL	COMMUNICATION SPEED	DISPLAY		OPTIONAL FEATURES	COST		COMMENTS
		TYPE	SIZE		BASIC*	OPTIONS	
Beeline Terminals Mini Bee 2	110-9600. Bits/Sec.	CRT	80 Char. X 25 Lines		$1,795		
DEC VT55	Up to 9600 Bits/Sec.	CRT (5 x 7 Dot Matrix Characters)	80 Char. X 24 Lines	Built-in Hardcopy Device	$2,495	$3,295 With Hardcopy	Standard: Character/Line Transmission, Line Editing, Cursor Control, Scroll Key, and Basic Graphics
IBM 3275	1200-7200 Bits/Sec.	CRT (12")	40 Char. X 25 Lines		$4,770		Full Screen can be Buffered For Block Transmission, Allowing Editing (Insert, Delete, and Erase)
Lear Siegler ADM-3	75-19,200 Bits/Sec.	CRT (12" With 5 x 7 Matrix Characters)	80 Char. X 12 Lines	24-Line Display, Numeric Keyboard	$995		
Teletype 33 KRS	110 Bits/Sec.	Impact Printing (Cylindrical Type Wheel)	72 Char./Line	Paper Tape Reader/Punch	$735		Heavy Duty Line Available (Model 35)
Teletype Model 40	110-4800 Bits/Sec.	CRT	80 Char. X 24 Lines	Storage For 2 Screen Images, Impact Printer	$3,525	$450 For Additional Storage, $1,850 For Printer	Standard: Full Cursor Control, Storage of full Screen, Full Editing and Many Special Features
Texas Instruments TI 743	110-300 Bits/Sec.	Thermal Hard-Copy (5 x 7 Dot Martix Characters)	80 Char./Line	Portable with Modem and Acoustic Coupler (745)	$1,395	TI 745 $1,995	This Model is one of TI's Silent 700 Series. The Line Ranges From the TI 743 to Programmable Models with Dual Tape Cassettes

* All Systems Will Require Modems To Communicate Over Voice-Grade Phone Lines. This Additional Device Will Cost Approximately $450.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

4. MICROCOMPUTER BASED SYSTEMS

Microcomputers represent the lowest cost approach for SEASAT data users to have local processing capabilities. Functionally, microcomputers can perform virtually all operations that larger computers can. Currently, however, they are lacking in processing speed. Also, support in the form of software and peripherals commensurate with the microcomputer's low price is not yet available in a large variety or quantity. As a minimum, microcomputer systems could carry on interactive communications with a central SEASAT data repository much like the terminals described in Section 3. In addition, they can enhance this interaction with the following capabilities:

1) Editing of user requests before transmission to the central site - users can enter a request and examine it to verify it is correct before transmission

2) Formatting of user requests into a standard coded form
   - Could allow users to use a variety of reference systems for geographical locations but all would automatically be converted to a standard for communications with the central site.
   - Could reduce amount of data to be transmitted.

3) Decoding of replies from the central site
   - Could convert a standard reference scheme to that of one currently being employed by the user.
   - Could greatly reduce amount of data which must be transmitted since the local processor could expand it for display purposes or directly store it to an intermediate media before processing.

4) Local manipulation of SEASAT data, such as searching or comparing locally stored data.

Limitations of microcomputer based systems lie mainly in their lack of speed and their customarily small memory and mass storage devices. Both limit their ability to perform complex computations or massive data manipulation without consuming much time and programming effort.

Of the categories of equipment discussed in this report, this one is currently experiencing the most rapid rate of evolution. Microprocessors themselves are relatively new and their potential is just now starting to be realized. New, significantly more powerful products are continually entering this market. The representative equipment described below is primarily from well established and reliable vendors. Competition in this field is sure to cause improvements in price and performance. In fact, a number of smaller vendors are currently offering microprocessor systems in kit or assembled form at extremely low prices compared to these major vendors. They are also offering increasingly complete lines of peripheral equipment. Of course, reliability and vendor support may prove real problems, and software development is left primarily to the users of these systems. They do, however, indicate the rush of such equipment we can expect in the near future. Many of the capabilities assigned to minicomputers in this report will no doubt become available in microcomputer systems over the next few years.

Following is a list of the equipment suggested for a microcomputer configuration, and some options which may be appropriate for particular applications.

1) Basic Systems

- Microprocessor for control and computational capability (requires programming language)

- Communications capability
- Local storage media-small volume such as floppy disk or cassette tape
- Small memory-approximately 16K words
- Control monitor (keyboard and display)

2) Options

- Printer
- Card Reader
- Graphics capability (would require CRT)

Two types of systems currently fall into this category. They are the packaged, "desk-top" units such as the IBM 5100 and the HP 9830, and those made of components centered around a basic processor, such as the PDP 11V03 or the GA 16/220. While the packaged units are easiest to set up and begin operation with, they are usually limited in memory size and potential for memory expansion. They are also generally limited in software and programming languages. In contrast, the component systems have nearly all the flexibility of full-sized computer systems. They are complete with operating systems, programming languages, assemblers, compilers, and a wide variety of peripherals.

One particularly unique piece of equipment included in this category is the Tektronix 4051. This device offers as standard, a microprocessor controlled graphics capability. Table 4-1 describes and prices the devices chosen as representative of this category. A wide variety of packaging and capabilities are included in this sample, and these unique features are noted in the table.

TABLE 4-1. MICROCOMPUTER BASED SYSTEMS

MANUFACTURER AND MODEL	COMMUNICATIONS SPEED	CONTROL MONITOR	WORD SIZE	MEMORY OPTIONS	LOCAL STORAGE OPTIONS	SOFTWARE AVAILABLE	OTHER AVAILABLE OPTIONS	BASIC SYSTEM COST*	COMMENTS
Data General Micro NOVA (Component System)	50 to 19,000 Bits/Sec.	ASR 33 Teletype	16 Bits	Up to 64K Bytes	Dual Floppy Disk Standard	Disk Operating System, FORTRAN Assmbler, Library	PROM Programmer	$12,285 (With 16K RAM)	The Basic System quoted is a Standard Packaged System Referred to as Their Development System
Digital Equipment PDP 11V03 (Component System)	50 to 9600 Bits/Sec.	LA 36 Keyboard Printer or CRT For Same Price	16 Bits	Up to 56K Bytes	Dual Floppy Disk Standard	Assembler, BASIC, Real-Time Operating System	FORTRAN ($700)	$12,225 (With 16K RAM)	
General Automation GA 16/220 (Component System)	Up to 9,600 Bits/Sec.	Carmel CRT (512 Char.)	16 Bits	Up to 128K Bytes in 8K Modules	Floppy Disks Disk Packs Magnetic Tape Cassette and Cartridge Tape	Assembler, FORTRAN, BASIC, Optional COBOL	Real Time Operating System ($1,500)	$16,350 (With 16K Memory)	Fully Compatible With The GA 16/440 (Minicomputer) and Can Use Same Software and Peripherals
Hewlet Packard HP 9830 (Packaged System)	Up to 1,800 Bits/Sec.	Built-in 32 Character Display	16 Bits	8 to 48K Bytes	Tape Cassette Standard Optional 4.8M Byte Mass Storage	BASIC, Operating System, Applications programs	CRT, Card Reader, Full Line of Printers, Plotter, Tape Reader	$10,500 (With 15,808 Bytes of RAM)	
IBM 5100 (Packaged System)	Up to 300 Bits/Sec.	Built-in CRT (5" diagonal-64 Char. X 16 Lines)	16 Bits (Plus 2 Parity Bits)	Up to 64K Bytes RAM Up to 190K Bytes ROM	Tape Cartridge Standard	BASIC, APL, Applications Programs		$8,975 (With 16K RAM)	
Tektronix 4051 (Packaged System)	110 to 2400 Bits/Sec.	Built-in CRT (11" diagonal Graphics CRT)	16 Bits	8 to 32K Bytes in 8K Modules	Tape Cartridge Standard	BASIC, Operating System, Graphics Commands	Graphics Hardcopy Unit, External Mag. Tape, Plotter, Joystick cursor Control	$9,145 (With 16K Memory)	The Outstanding Feature of This Unit is its Graphical Capabilities

* The Basic System Includes:
- Communications Capability
- Local Mass Storage (Floppy Disk or Magnetic Tape)
- Small Memory
- Control Monitor (Keyboard and Display)

5. SMALL MINICOMPUTER BASED SYSTEMS

Minicomputer systems can offer SEASAT data users graphical data entry and display, as well as provide on-site processing capability superior to that of most microprocessor systems. Most graphical systems have been developed for minicomputer and larger systems, and can benefit greatly from memories and mass storage devices larger than those generally associated with microprocessors. The processing advantage of minicomputers over microcomputers is a result of both their generally higher execution speed and their currently more sophisticated software. They provide users with higher level languages (i.e. FORTRAN and ALGOL), file handling software, and some even have generalized Data Base Management Systems (DBMS's). Such support helps make machine operation of less a concern to users. The minicomputer based systems described below could handle all the functions described in the previous section on microprocessors (i.e. communicate with the central site, format requests, decode replies, and provide local processing of SEASAT data), plus the following:

1) graphical display of SEASAT data and maps.

   - users could reference points on their display which would automatically be converted to a standard reference for communications with the central site

   - geographical overlays could be stored on-site and communication with the central site would only have to be concerned with changes or additions to these overlays (great reduction in required transmission)

   - data to generate SEASAT maps could be transmitted on a standard media through the mail for local use and to add to local libraries of data

2) More computational capability due to faster processors, more storage, and more extensive software.

The systems included in this category have a great deal of processing potential, but in order to limit their price, they have been restricted to 16-bit word machines with only minimal memory and peripherals. They can handle graphics and are adequate for most kinds of stand-alone processing, but could prove awkward for large data manipulation problems.

The following lists of basic system and optional components specify the types of systems included in this category:

1) Basic Systems

- 16-bit Minicomputer
- communications capability to interface with central site over voice-grade phone lines
- 10 million byte disk
- magnetic tape unit
- medium sized memory, approximately 32K words (64 K bytes)
- slow speed printer and card reader
- graphics capability with CRT

2) Options

- hardcopy for graphics
- higher speed I/O devices
- more memory and mass storage

Table 5-1 summarizes some of the equipment currently available in this category. As technology advances and microprocessors exert more pressure to take over the tasks of small minicomputers, we can expect competition in this area to become even keener. These systems can provide an excellent starting point to be upgraded and augmented into ultimately very powerful systems.

TABLE 5-1 SMALL MINICOMPUTER BASED SYSTEMS

MANUFACTURER AND MODEL	MEMORY OPTIONS	LOCAL STORAGE OPTIONS	SOFTWARE AVAILABLE	GRAPHICS SYSTEM	PRINTER OPTIONS	CARD READER OPTIONS	BASIC SYSTEM COST*	COMMENTS
Data General Nova 3/12	8 to 128K words	Diskettes Cassettes Mag. Tape Disk Packs	FORTRAN ALGOL, BASIC, Assembler	D. G. 6012	80 to 600 lpm	150 to 1000 cards/min.	$59,000	
Digital Equipment PDP 11/34	8 to 128K Words (4K for Systems)	Disk Packs Mag. Tape Cassette Tape	FORTRAN, BASIC, PDL, Plotter Package, Library	VS 60 Graphic Display System	60 to 1200 lpm.	300 to 1200 cards/min.	$87,065	Hardware floating point optional, user microprogramming, Memory Management Available
General Automation GA 16/440	Up To One Million Words	Disk Packs Mag. Tape Cassette or Cartridge Tape	FORTRAN, BASIC, Assembler	Tektronix 4002	200 to 1000 lpm.	300 to 1000 cards/min.	$59,350	Hardware memory mapping, multi-user BASIC Available
Honeywell Level 6	8 to 16K Words	Cartridge Disks	Assembler, FORTRAN	Honeywell 606	240 to 600 lpm.	300 to 500 cards/min.	Not Available at This Time	Hardware multiply/divide standard
Interdata 7/16	8 to 64K Words	Disk Packs (Up to 300M Bytes) Mag. Tape	Assembler, FORTRAN,	Tektronix 4010	200 to 600 lpm.	400 to 1000 cards/min.	$70,000	
Modcomp II (26)	Up to 64K Words	Disk Packs Mag. Tape Floppy Disks	Assembler, FORTRAN, BASIC	Tektronix 4010	300 to 600 lpm.	300 to 1000 cards/min.	$72,546	
Prime 300	Up to 256K Words	Disk Cartridge Disk Packs Mag. Tape	FORTRAN, Assembler, BASIC, Library	Tektronix 4014	60 to 1100 lpm.	300 to 1000 cards/min.	$74,400	

* The Basic System Includes:

- 16-bit Minicomputer
- Communication Capability
- 10 Million Byte Disk
- Magnetic Tape Unit
- 32K Words of Memory
- Graphics Capability with CRT
- Slow Printer (approximately 250 lpm.)
- Slow Card Reader (approximately 300 cards/min.)

One device which does not easily fit into Table 5-1 is the printer/plotter. Generally, employing a thermal process of matrix dot printing, these devices can function both as medium to high speed printers and as graphical hardcopy units. Their cost rivals that of either of these devices, making them an attractive alternative. Operating costs, however, are higher and must be considered. Manufacturers of such devices include Versatec, Varian, and Gould.

6. LARGE MINICOMPUTER BASED SYSTEMS

Whereas the minicomputer systems described in the previous Section were limited to minimal configurations to minimize their costs, the minicomputer systems included in this Section are designed primarily for computing power. These systems are powerful enough for massive data manipulation problems. In order to take advantage of their computing power, most have software available to support communications with networks of processors or terminals. All have the capability to execute batch jobs while performing real-time processes (such as interactive communications). Again, graphics capability is a primary feature of these systems, particularly for SEASAT applications. The following are the primary capabilities of the large minicomputers described in this Section:

1) Intelligent communication with a central computer or network where requests can be coded to a standard form and replies can be decoded in terms of the users points of reference
2) Extensive graphical display and processing of SEASAT data (same as for small minicomputer system, but able to handle larger amounts of data more efficiently)
3) Able to support multiple users and could even be configured to support a network of users with their own terminals. It could be a local or regional center for data from SEASAT and other projects
4) Its stand-alone potential would make it appropriate for a wide variety of local applications
5) The system could support color graphics

Of course, such processing power is not inexpensive. These systems would be appropriate for shared facilities where total cost is defrayed over many users. These users must balance having the power of such a system part-time against having a smaller dedicated system. The following list specifies the minimum and optional components of systems in this category:

1) Basic System

- 32-bit minicomputer
- Communications capability to interface with central site over voice-grade phone lines
- 50 million byte disk
- Magnetic tape unit
- Large memory, at least 64K words (256K bytes)
- Graphics capability with CRT
- Hardcopy for graphics
- Medium to high speed printer and card reader

2) Options

- Color graphics
- Higher speed I/O devices
- More memory and mass storage

The 32-bit word length specified above insures that arithmetic accuracy is adequate for almost any applications.

Table 6-1 presents those systems chosen as representative of this category. Just as for small minicomputer systems, the printer/plotter does not easily fit into the table, but should be considered as a possible part of any of these systems. Currently, the large minicomputer field is not nearly as crowded as the small minicomputer field. This, however, is expected to change as minicomputer companies become

TABLE 6-1 LARGE MINICOMPUTER BASED SYSTEMS

MANUFACTURER AND MODEL	MEMORY OPTIONS	LOCAL STORAGE OPTIONS	SOFTWARE AVAILABLE	GRAPHICS SYSTEM	PRINTER OPTIONS	CARD READER OPTIONS	GRAPHIC HARDCOPY DEVICE	BASIC SYSTEM COST*	COMMENTS
Interdata 8/32	120K To 1M Bytes	Cartridge Disk, Mag. Tape, Cassette Tape	FORTRAN, Assembler, BASIC	Tektronix 4010	200 To 600 lpm.	400 To 1000 Card/Min.	Tektronix	$138,000	Will Support OEM Color Graphics True 32-Bit Machine
Modcomp IV (25)	64K To 1M Bytes	Moving Head Disk, Mag. Tape	Assembler, FORTRAN, BASIC, Application Library	Tektronix 4010	300 To 600 lpm.	300 To 1000 Card/Min.	Tektronix	$153,000	Uses 16-Bit Memory and Double Fetches for Each Word
Prime 400	Up to 8M Bytes	Cartridge Disk, Disk Pack, Mag. Tape	Assembler, FORTRAN, COBOL, BASIC	Tektronix 4014	300 To 1100 lpm.	300 To 1000 Card/Min.	Tektronix	$137,000	Virtual Memory Capability, Requires 16-Bit Memory, Two Parallel Fetches for each Word (Memory Bus is 32-Bits Wide)
SEL 32/55	32K To 1M Bytes	Cartridge Disk, Mag. Tape	FORTRAN, Assembler, Application Library	Tektronix 4631	300 To 600 lpm.	285 To 1000 Card/Min.	Tektronix	$148,400	Virtual Memory Capability True 32-Bit Machine

* The Basic System Includes:

- 32-Bit Minicomputer
- Communications Capability
- 50 Million Byte Disk
- Magnetic Tape Unit
- 64K Words (256K Bytes) of Memory
- Graphics Capability with CRT
- Hardcopy for Graphics
- Medium Speed Printer (approximately 600 lpm.)
- Medium Speed Card Reader (approximately 1000 Card/Min.)

more sophisticated and offer more complete systems which can compete with larger computers. Although the systems described in this section are mini's, many are as powerful as the full-sized computers of a few years ago and at a fraction of the cost.

7. EQUIPMENT PROJECTIONS THROUGH 1985

User requirements and technological advances will be the forces that shape the future of computing systems. Of course, these forces are interrelated in that technological advances help to stimulate new user requirements, while user requirements often prompt research in particular areas. Some of the expected changes that can be attributed to technological advances follow:

1) Costs will drop dramatically for equipment to provide a given level of capability. This will be a result of the historical cost trends for electronic devices.

2) Mechanical components will be increasingly replaced by electrical components. This is most applicable to the peripheral area where we expect more CRT's, electronic mass storage, and electronic print mechanisms (the transporting of physical media, such as printer paper of course must remain mechanical, but even the quantity of this operation could be significantly reduced by using such media as microfilm). The reason for this trend will be the continuing drop in electrical component prices while mechanical devices have basically reached a price plateau.

3) New architectures will be developed and perfected to take advantage of the evolving relative cost relationships between components. For example, as processor costs drop in relation to peripheral prices, it becomes cost effective to develop multiprocessor configurations with common peripherals.

Some less technological responses to user requirements can also be predicted. As cost go down and corresponding capabilities increase, computing will gain even further acceptance accompanied by an increasing variety in the sophistication and financial resources of users. The following represent some expected responses:

1) More complete system support will become available, particularly for small systems (currently the small minis and micros). Historically these small systems were used by specialists for specific applications. As the power of these systems increase and their costs decrease, they will be attractive for a more general group of users. In order to serve this potentially huge market, vendors will be prompted to offer added service so that machine operations become of less concern to their users.

2) Standard software packages will become more prevalent. In general, equipment prices are dropping while software development costs are not. Software packages will allow the defraying of developmental costs over many users. As such packaging becomes more prevalent, we can expect to see more competition and improved products.

3) Computer systems will become more modular so that they can be easily configured, assembled, and upgraded. Peripherals in particular are becoming available from such a variety of vendors, that these vendors must strive for compatibility to insure their very future. Hopefully, we will see more interface standardization. Adding dedicated processors to peripherals and to peripheral handlers will provide added flexibility in interfacing equipment from various vendors.

The following subsection examines the individual categories of equipment discussed in Sections 3 through 6 in terms of expected changes in capabilities or configurations, while the final subsection presents projection of system costs through 1985.

7.1 CAPABILITIES BY EQUIPMENT CATEGORIES

This subsection examines each of the equipment categories discussed in Section 3 through 6, and describes the type of equipment expected to be available in the future to meet SEASAT data processing requirements.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

7.1.1 Telecommunications Terminals

Telecommunication terminals are appropriate when a single, central data base must be accessed or when local processing capabilities are not adequate. Many SEASAT data users will be interested in only data for their particular region, which could be made available directly to them at their own sites. Currently, however, on-site processing of SEASAT data requires equipment considerably more expensive then telecommunications terminals. Communications costs are not expected to drop as significantly as costs for computers. As costs of small computing systems approach costs incurred in using telecommunications terminals, SEASAT users will switch to local processing capabilities, if even only to perform minimal data coding and decoding.

As logical devices (i.e. microprocessors) drop in price, they will increasingly become an intergal part of most teleprocessing devices. They will add flexibility to device control (various interfaces could be provided by different control programs for the processor) and provide for easier use (on-site editing and user-specific commands could easily be supported). Similarly, as storage media becomes less expensive, it will be more practical to store intermediate data on site, perhaps for editing, reference purposes, or block transfers of data. Currently, many telecommunications terminals still employ impact printing techniques. Price graphs show that while these devices have experienced minor cost reductions, CRT terminals are in midst of a steep price decline, with some already in the same price range as the impact devices. Basicly electronic devices,

CRT's should continue this price decline, replacing impact terminals where hardcopy output is not necessary. In fact, various electrostatic printing techniques are becoming increasingly popular both for reasons of price and reliability. It is expected that such techniques employing standard paper will be perfected and displace the impact devices. A likely configuration will be a keyboard, CRT, and optional electrostatic hardcopy device. Such configurations will offer reliability, quiet operation, and a price below current telecommunication terminals.

7.1.2 Microcomputer Systems

In the previous sections, microcomputers represented the least expensive and least powerful level of computing appropriate for individual SEASAT data users. The LSI components which previously delineated the category of microprocessors are now finding their way into even the largest computers. By 1985, microprocessors will have surpassed the performance available in todays minis and will be components of all sizes of computers. Thus, the category of 1985 equipment safisying the requirement for a small system, low in cost and responsible for a single application at a time will be refered to as a small personal computer (without reference to technologies used). Similar to the microcomputer systems discussed in Section 4, SEASAT data users would require on-site storage, a processing language, and communications capabilities.

The small personal computers of 1985 will be a response to a demand for such capability from a wide variety of users. As such, ease of operation will be a key improvement over todays small systems. Higher level languages, special purpose languages, and a variety of packaged software will be available. Processing speed, even on these low priced systems is likely

to increase, but overhead from operating systems providing services to the users will impact throughput. In order to increase speed and flexibility, even these basic systems are likely to have distributed processing with individual processors dedicated to particular applications (e.g. user interface, peripheral control, and arithmetic computation). Some of the mass storage on these systems is likely to be electronic in nature, but regardless of technology, total capacity will be much increased over the present media. Higher packing densities, improved media, and new storage approaches will lead to this improvement with a certain decrease in cost. Small CRT's integrated into these systems will be very popular as their prices drop well below that of current terminals. Many may also have hardcopy devices which can be used on demand to generate a permanent image of the CRT screen. It will be important to SEASAT data users that this sytem is able to efficiently acquire SEASAT data. This may be through high speed communications relaying data to mass storage. A simpler approach, however, would be to allow for some removable mass storage media. Data in standard format could be supplied to users on this media for immediate inclusion into their systems.

By 1985, graphical reporting will be very popular. By then, required equipment will be inexpensive and it will be practical to dedicate processors to generating graphical output and to simplifying user interface. In fact, it is reasonable that by 1985, a small personal computer with a CRT monitor as described above could be augmeneted with a graphics module for very little cost. The increased processing load would be primarily handled by the graphics module itself.

7.1.3 Small Minicomputer Systems

As indicated previously, the destinction between mini and microcomputers will become blured in the future. Further, the small minicomputer systems described in section 5 were configured for SEASAT data users requiring graphics capabilities, and by 1985, intelligent graphics terminals capable of interfacing with even very small computers will be available. This category of equipment will continue to respond to the demand by SEASAT data users for relatively powerful, dedicated systems. With such a system, a user can access the system and any of its stored data on demand. By adding special processors, even the traditional "number-crunching" types of problems could be handled by relatively inexpensive machines (affordable for individual users). For projection purposes, we will consider this class to be those computer larger than the minimal systems described in the previous section but without facilities or software for supporting multiple users simultaneously.

As steps are taken to make these computer systems easier to use, operating systems will be refined and user communications simplified. Data base management features will become available with virtual storage so that users will be able to easily reference very large data bases. Graphics usage will expand as intelligence in graphics terminals allow extended capabilities. Price reductions in storage media will make colored graphics feasible for individual applications. Considering the usefulness of multi-color geographic representations of SEASAT data, we expect wide application of colored graphics by these users.

The basic equipment configuration of these future systems will be similar to that of the small minicomputers of section 5. Much larger electronic mass storage devices will no doubt be incorporated into these systems. It is likely that mechanical printers will be replaced by some type of electronic printing mechanism which will also be able to produce graphics output and images of the CRT screen. It is also possible that card readers will be replaced by magnetic data input devices in many instances.

7.1.4 Large Minicomputer Systems

The large minicomputer category will be very hard to delineate in the future. Their current trend is toward general purpose capabilities and they are making serious inroads into the full-sized computer field. For projection purposes, we will consider this category to evolve to a class of multi-user, general purpose computers with capabilities anywhere from their current level to that of our more sophisticated current computer systems.

Notable advances expected in this category include:

1) Very large, on-line mass storage devices (electronic in nature)
2) Virtual memory and Data Base Management Systems
3) Multiple processors, allowing these systems to service large numbers of users concurrently and to fully utilize all components
4) Shift from punched card input to magnetic media data input
5) Wide use of electronic printing devices also capable of graphics
6) Colored graphics on CRT
7) Colored hardcopy of graphics output, perhaps directly to film rather than paper
8) Advanced microfilm output.

These systems will continue to be most appropriate for groups of SEASAT data users able to share the use and cost of such systems.

As shown above, we expect both refinements of existing capabilities as well the addition of others. This usually is the trend for general purpose equipment. Currently, large minicomputer systems are just building their reputations as complete and well supported systems. Over the next few years, we will see major efforts by vendors to supply support and software to make them more attractive to the general computer using community. As in all levels of computing, use of many processors to control devices and provide computing power seems inevitable, just as the move from mechanical to electronic peripherals.

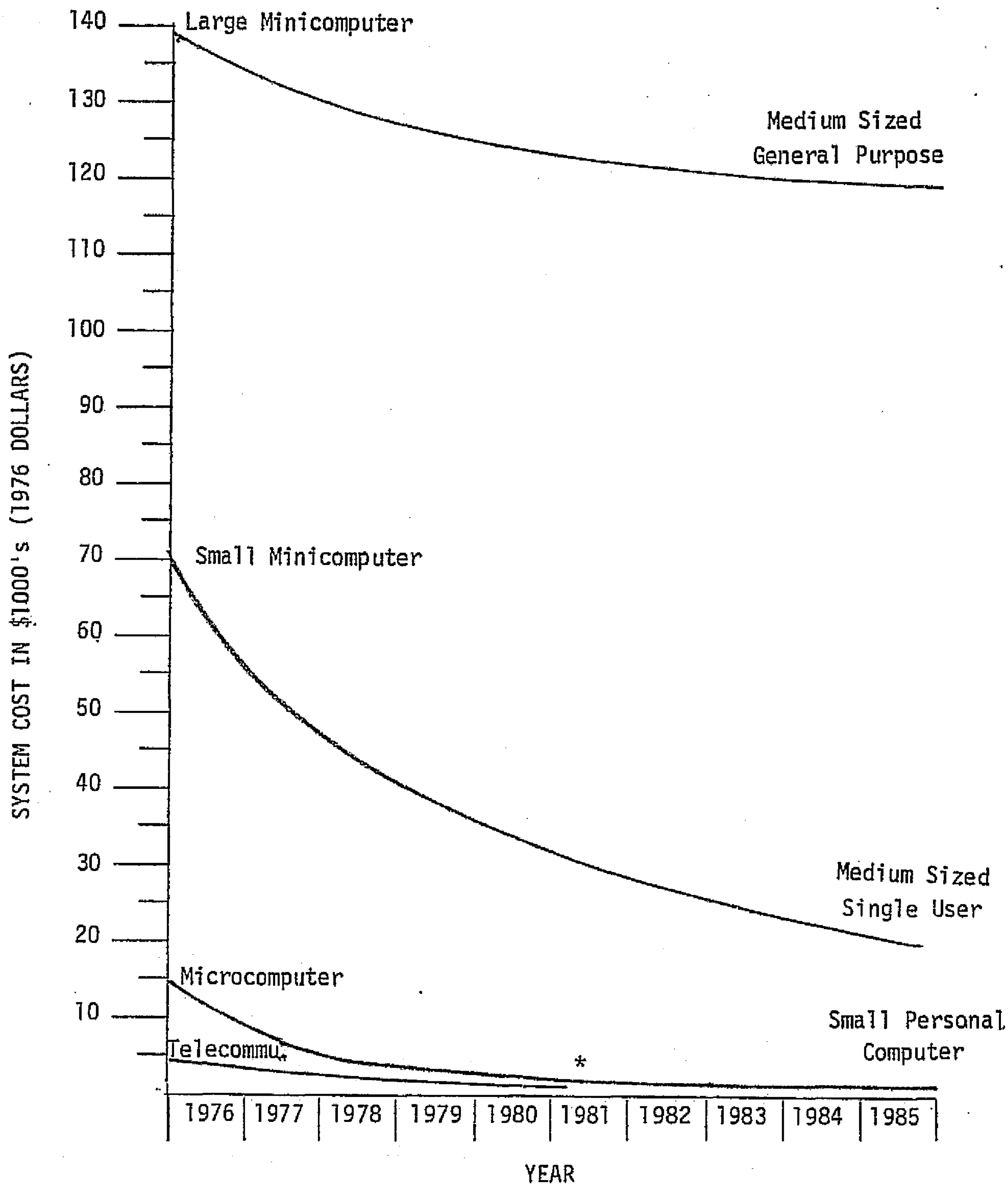
7.2 SYSTEM COST PROJECTIONS

The cost projections presented in this subsection are based on recent price trends for computer equipment and the historic price curves for other electronic devices. These projections are for complete processing systems, as described in previous subsection, rather than for individual components (except the category of telecommunication terminals). All cost projection are in terms of 1976 dollars. Table 7-1 summarizes how the naming and price ranges of these categories are expected to evolve by 1985. This evolution of names is based on the uncertain future of such terms as micro and mini. Figure 7-1 presents cost curves through 1985, showing continuous cost projections for average systems in each category.

CURRENT EQUIPMENT		1985 EQUIPMENT	
CATEGORY	SYSTEM COST	CATEGORY	SYSTEM COST
Large Minicomputer Based	$137,000 to $153,000	Medium-Sized General Purpose	$100,000 And Up
Small Minicomputer Based	$ 50,000 to $ 87,000	Medium-Sized Single User	$ 10,000 to $ 30,000
Microcomputer Based	$ 9,000 to $ 16,000	Small Personal Computer	$ 1,000 to $ 5,000
Telecommunications Only	$ 1,000 to $ 4,800	Not Applicable For SEASAT Users	———

TABLE 7-1 SUMMARY OF EQUIPMENT CATEGORIES AND COSTS

The category of telecommunications equipment is not carried through 1985 on either Table 7-1 or Figure 7-1. This is not to imply such equipment will not exist by then, because certain applications surely will continue to require it (where a central data base is required or where local processing is impractical). In the near future, however, it is expected that the cost of small personal computers will approach that of the telecommunications terminals and it will become more appropriate for even very small SEASAT data users to acquire such systems rather than simple terminals. These systems will be able to accept SEASAT data, perform searches and comparisions of the data, and probably even display it graphically.



*As the prices of small personal systems approach those of telecommunications terminals, most SEASAT users are expected to choose the small computing systems over the simple terminals.

FIGURE 7-1 COST PROJECTION

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

As illustrated in Figure 7-1 the relative slopes of the various cost curves are not the same. The systems configured for single users of specific applications (small mini's and micro's) are shown decreasing in relative cost much faster than the larger, more general purpose equipment (large mini's and the telecommunication terminals). Historically, equipment designed for specific applications takes advantage of technological advances primarily by lowering costs rather than by significantly increasing capabilities. On the other hand, users of general-purpose equipment tend to acquire and use as much equipment as they can afford. As new capabilities become available and component costs drop, they will often maintain their level of investment or even raise it to acquire significantly improved equipment. Of course, the cost for a fixed set of computer capabilities (if even definable) will drop dramatically over this time period, but in this general-purpose category, we are expecting these drops to be offset by additions of capabilities and new components. Cost per processor may drop substantially, but each system may employ many processors along with expanded vendor and software support. The telecommunications terminals show a gradual cost reduction because of their dependence upon mechanical components and because they are further along in their evolutionary cycle than the other devices discussed.

PART 2

DMS 1100 TEST AND EVALUATION

Teledyne Brown Engineering (TBE) recently exercised the Data Base Management System DMS 1100 as installed at NASA's Marshall Space Flight Center. This exercise was designed primarily to obtain quantitative measurements of the implementation and operating overhead introduced by the use of such systems, in terms of added manual and computing efforts. Part 2 of this report describes the tests performed, the results of these tests, TBE's evaluation of the results, and recommendations to potential Data Base Management System users to assist them in deciding when such a system is appropriate for their needs. Specific comments relating to DMS 1100 as a representative system are included with these recommendations.

1. INTRODUCTION

Data Base Management Systems (DBMS's) have been hailed as the major software advancement of the 70's. In concept, they certainly hold much promise in terms of centralizing data management responsibilities, reducing data redundancy, and providing users with more flexibile data access. In terms of product life, DBMS's are still in their formative years and can expect many refinements in the next few years. Standardization, ease of use, and more efficient operation are all areas which should see improvement.

Many reports have been produced comparing the availability of particular features in various DBMS's. Frequently, these reports also assign weighting factors to determine those best suited for a particular application. Few potential DBMS users, however, have actually worked with such a system. It was the purpose of this study to evaluate a representative, well-established DBMS by actually exercising it with typical data access operations. A benchmark set of operations were developed to obtain quantitative measures of the operating overhead introduced by the DBMS. Emphasis also was placed on determining the level of effort and time required for data processing personnel academically familiar with DBMS approaches to initiate and use a typical DBMS.

The DBMS examined by this study was UNIVAC's DMS 1100, as implemented on a UNIVAC 1108 at NASA's Marshall Space Flight Center. DMS 1100 was selected because: it represents a major vendor's DBMS; it is operational at a number of sites; it conforms to the Conference of Data System Languages (CODASYL) recommendations for DBMS's (current moves toward standardization appear to be in this direction with most new DBMS's also conforming to these recommendations); and it was installed and available for exercise. Standard COBOL file management software was used as a point of reference for comparison purposes.

The following Section describes the tests performed and presents the quantitative results. Section 3 analyzes these results and their causes while Section 4 presents recommendations on when and how to use a DBMS with particular reference to DMS 1100.

2. TEST PROCEDURE AND RESULTS

This test of DMS 1100 was intended to provide quantitative measures of the overhead (both man and machine) introduced by the use of a typical DBMS rather than the normally available system utilities for data management. The environment and applications tested were chosen to be representative of those for many of the organizations now contempleting implementation and use of a DBMS. The following subsection describes in detail the tests performed, while subsection 2.2 presents the quantitative results of this effort.

2.1 TEST DESCRIPTION

There were actually two parallel parts to this test. One task was to exercise DMS 1100, while another was to perform corresponding benchmark tests with COBOL using system-standard file handling software. The following subsections describe the operating environment, file structures, and operations used for both parts of this test.

2.1.1 Environment

The computer environment for this test was a UNIVAC 1108 installed at NASA's Marshall Space Flight Center. This installation operates under the EXEC 8 operating system and jobs were submitted in a batch mode. Due to their heavy computing load and the low priority assigned to large jobs such as these data base tests, we were able to average only one batch of computer runs per day. Local memory and disk space constraints did not interfere with the execution of the tests.

Version 5A of DMS 1100 was used for this test since it was the most recent release available. A single thread configuration was used, in which each user of the system has to incorporate his own personel version of the DMS 1100 software into his program. Multi-thread configurations also exist, where a single copy of the DMS 1100 software resides at the operating system level and may be accessed by multiple users. Overhead for a single program to use DMS 1100 should be

similar for either configuration. Only in multiple concurrent user situations should the multi-thread configuration reduce total memory requirements. At the time of these tests, the multi-thread version was not available for our use.

In order to use DMS 1100, users must first design and create their data base. The data base design in itself is a major project in that requirements for the data must be established and the data base design optimized for efficiency and ease of use. The data base is defined using a special Data Definition Language (DDL) which is passed through a DDL Processor to produce an Object Schema. This Object Schema describes the structure of the data files for the computer and is referenced during processing of file manipulation commands. These commands are written using the Data Manipulation Language (DML). The DML is not an independent language, but rather uses COBOL as a host language (Version 5A of DMS 1100 uses UNIVAC's Fieldata COBOL). These DML commands are imbedded into COBOL programs and pre-processed by DMS 1100 software which translates them into COBOL compatible calls to the DMS 1100 Data Manipulation Routines (DMR's).

For benchmark purposes, standard file manipulation facilities available in UNIVAC's Fieldata COBOL were used to establish reference statistics. The Indexed Sequential File Management System (ISFMS) software package was included in the version of COBOL used to provide added flexibility in data access.

2.1.2 Files

The files of data used for this test were designed to be representative of the types of files in many large data management systems. In fact, they were modeled after an existing inventory control and reporting system. These files varied in size and record length as follows:

FILE NUMBER	NUMBER OF RECORDS	RECORD LENGTH IN WORDS OF COMPUTER MEMORY
1	700 Records	7 (42 Characters)
2	16,000 Records	21 (126 Characters)
3	23,000 Records	14 (84 Characters)

Each file was first created as a COBOL file and then a corresponding DMS 1100 file was constructed using the DML. All test files were created once, stored on magnetic tape, and loaded to disk before individual test runs.

2.1.3 Operations on Files

The operations performed and timed for this test were those that occur frequently in updating and accessing a large data base and included operations requiring random access to stored data. Table 2-1 describes the data manipulation activities tested and the particular operations used in DMS 1100 and Standard COBOL with the ISFMS to accomplish these activities.

In addition to the data manipulation activities described above, it was our original intention to examine the operational attributes of DMS 1100 in an expanding file environment. It is a well known characteristic of the standard indexed sequential file handling package, ISFMS, that as records are added to a file, causing a large number of entries in the overflow area, the time required to fetch or store new records increases dramatically. This increase is due to additional time required to search the sequential overflow area and can be alleviated only by reorganizing the file. We had intended to investigate the behavior of DMS 1100 in a similar environment. However, in examining the DMS 1100 documentation, we discovered that DMS 1100 uses the standard file handler and contains no special provisions for automatic file reorganization. We expect DMS 1100, therefore, to exhibit the same increase in overhead as ISFMS in an expanding file environment.

2.2 QUANTITATIVE TEST RESULTS

DMS 1100 required significantly more resources to operate than the standard COBOL file management utilities. The programming effort

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

required to use DMS 1100 consumed many times that required by the standard COBOL, and computer execution time for those parts of the programs attributed to data access and management were tripled by DMS 1100. Timings for the operations tested were acquired using a system utility which allowed measurement of elapsed CPU time at any point in a programs execution. The timing results obtained for DMS 1100 and for the reference COBOL environment are summarized in Table 2-2. The following Section analyzes the results of these tests.

DESIRED ACTIVITY	DMS 1100 COMMAND	COBOL ISFMS COMMAND
• Locate A Record With Known Key	Random "FETCH"	Random Read
• Sequential Scan Through File	"FETCH NEXT RECORD"	Sequentially Read Next Record
• Insert New Records	Random "STORE"	Random Write
• Operations Normally Associated With Updating A Record (Random Read, Rewrite The Record, And Reread The Record As A Check)	"FETCH", "MODIFY", AND "FETCH"	Random Read, Rewrite, And Reread Of The Same Record
• Delete A Record	Random "DELETION"	Random Delete

TABLE 2-1 OPERATIONS PERFORMED FOR TEST

OPERATION	MEAN CPU TIME REQUIRED PER OPERATION (IN MILLISECONDS)		MULTIPLE OF COBOL ISFMS TIME REQUIRED BY DMS 1100
	DMS 1100	COBOL ISFMS	
Random Read	8.50	2.95	2.9
Sequential Read	2.20	1.00	2.2
Insert New Record (Random)	11.07	4.60	2.2
Update Record (Read, Rewrite, and Reread The Same Record)	13.60	3.60	3.8
Delete Record (Random)	9.90	2.78	3.6
Average	8.85	2.99	3.0

TABLE 2-2 QUANTITATIVE RESULTS OF TESTED OPERATIONS

3. ANALYSIS OF TEST RESULTS

These tests showed an unexpectedly high overhead in the use of both manual and machine resources required to operate this representative DBMS. The following subsections comment on these two aspects of overhead and note their causes.

3.1 EASE OF USE

Ultimately, DBMS's should reduce the amount of programming effort necessary to use large data files by providing easier, more powerful data manipulation capabilities than are currently available in standard systems. Furthermore, a unified data base should centralize responsibility, optimize the data structure, and ultimately result in long term savings. Such a refinement of DBMS's was not evidenced by this test. The potential for such personnel savings no doubt exists in DMS 1100, and probably becomes tangible after users become very familiar with the system. Our test, however, which emphasized initiating the use of such a system, pointed out that the user interface portions of the system still need much refinement.

The easiest way for a person to learn to operate a new system is to have the assistance of someone who has already used and is familiar with the system. Unfortunately, that body of personnel is relatively small for DBMS's. Thus, most new users will have to "go it alone" to some extent. Such was the case for this test. We were able to locate few current users of DMS 1100 at MSFC and the operational applications did not use all of the facilities to be tested. Faced with such a situation, system documentation and program "run-time" diagnostics became the primary sources of information and assistance. In general, both of these sources were found lacking in completeness and clarity. This is not surprising for such a major, relatively recent product, but does point out a weakness expected in most current DBMS's.

3.1.1 Documentation

In using DMS 1100, the following documentation was used:

DMS 1100 Schema Definition -(UNIVAC Pub. No. UP-7907, Rev. 2)

DMS 1100 ANS COBOL (FIELDATA) Data Manipulation Language - (UNIVAC Pub. No. UP-7908, Rev. 1A)

ANS COBOL (FIELDATA) -(UNIVAC Pub. No. UP-7845, Rev.C)

Programmer Procedures Manual -(MSFC Publication)

CODASYL Data Description Language - Journel of Development, June, 1973 - (U.S. Dept. of Commerce Publication - NBS Handbook 113)

Date, C. J., An Introduction to Database Systems, Addison Wesley Publishing Company, 1975.

The first three manuals listed above represent the UNIVAC documentation appropriate for DMS 1100 users. This documentation assumes a great deal about a potential user's background in database management systems. Some critical terms were poorly defined or not defined at all and an overview of system operation was not included. Also lacking were hints on how to structure individual databases to take advantage of the inherent internal structure of DMS 1100. Such information is valuable but difficult to determine, particularly by new users. Most emphasis in the UNIVAC documentation was placed on describing the construction of valid commands (syntax), while much less effort was given to completely specifying the action caused by these commands (semantics). In both cases, some omissions and ambiguities were noted, but primarily in the area of semantics. The other documents referenced in the above list had to be consulted to augment the UNIVAC manuals, and even then, not all questions which arose could be answered.

3.1.2 Program Diagnostics

Poor run-time diagnostics have been a traditional complaint of programmers, and new, large systems are generally the worst offenders. It must be appreciated that the larger and more complex a system is, the more

critical are these diagnostics since it is particularly difficult for users to comprehend these systems. Some of the problems encountered with DMS 1100 diagnostics are:

- The DDL Processor failed to detect some syntax errors
- Some of the error messages produced by the DDL processor were ambiguous
- Some of the run time diagnostics were totally inappropriate for the commands they referenced
- Some errors during file generation were not reported

Such problems are serious, particularly for a new user unfamiliar with the system, and even more so when they represent the only programmer assistance available (no experienced users for consultation). These difficulties significantly added to the DMS 1100 programming time required in this test and similar experiences are expected for other new users of DMS 1100.

3.1.3 Comments on System Design

The syntax of the DDL is very similar to COBOL, but in many instances, punctuation rules are different. Although the DDL commands are assembled to form an independent program which builds the object schema, COBOL serves as a host language for other DMS 1100 operations. Therefore, it seems reasonable that the DDL should be as similar to COBOL as possible. Further, it appeared that this punctuation was in fact unnecessary, but improper use of it still caused disastrous results. This design feature contributed to unnecessary difficulty in initializing DMS 1100 files.

3.2 MACHINE RESOURCES

Considering the additional capabilities and services provided by DBMS's over standard file management software, it is to be expected that they will require added machine resources. The fact that CPU time required for file related processing with DMS 1100 was three times that for standard COBOL with the ISMFS, was surprising. If, for a particular application the capabilities offered by DMS 1100 are necessary, or if savings in programming time can be established, such a price may not be unreasonable. Such an analysis, however, requires the context of a particular application, and therefore is not appropriate here. Instead, we can examine the source of this overhead.

The added processing time and data storage facilities generally required by DBMS's can be attributed to:

- maintenance of intricate data structures and presenting this data to users as if the data were stored in a form defined by the user
- logging and access checks to maintain data security in a multi-user environment
- activity logging to provide back-up protection.

All of these activities are likely to add processing and storage overhead, but the effect on storage requirements of the first (maintaining data structures) is particularly dependent upon the environment and the data structure that would be used without a DBMS. Given a single file, conversion from a standard environment to a DBMS environment may require up to three times the storage space. This will vary depending on the system and the interrelationships inherent in the data. When multiple files are combined into a DBMS, however, the change in storage requirements will also be affected by the reduction in data redundance that was present in the original files. In fact, there are cases of conversions to DBMS's where the increase in storage required per data element was totally offset by the reduction in the number of unique data elements.

A potentially serious drawback to the use of DBMS's is that these added computer requirements are often significant enough to actually necessitate added equipment. Computer facilities traditionally operate at near capacity, and the addition of a large DBMS data base often means upgrading the CPU and disk subsystem.

4. RECOMMENDATIONS

The conversion to and use of a DBMS is a major step and must be carefully considered. DBMS's are by no means a cure-all to data management problems and in certain cases they are not even appropriate at this time. The following subsection presents guidelines to help potential DBMS users decide when the use of a DBMS is appropriate. The final subsection addresses recommendations on the use of DMS 1100 as a representative DBMS.

4.1 WHEN IS A DBMS APPROPRIATE

In determining if a DBMS is appropriate, a potential user should first obtain a clear picture of the benefits offered by DBMS's over those available with standard file management features. If these benefits would be significant for the projected applications, then individual DBMS's should be investigated in terms of their particular features, acquisition costs, and operational costs.

The major potential advantages generally offered by DBSM's are:

- Responsibility for acquisition, update, and maintenance of data is centralized.
- Sharing of common data among multiple users is simplified since each user, internal to his own program, can define an appropriate structure for the data.
- Sophisticated data access techniques (i.e., indexed sequential with multiple keys) are available.
- Data Security is provided by DBMS's with various levels of access allowed individual users (in some systems down to the data element level).

These advantages are most appropriate for large data systems with multiple users and much common data, some of it priviledged or for restricted usage (such as a company-wide management information system including personnel

records and salary information). Probably the driving force for developing DBMS's was the elimination of redundancy in data bases and the centralization of control over data. Data security features developed out of necessity for privacy in this environment. Often these advantages can actually reduce programming overhead by centralizing data management responsibilities and providing sophisticated data access procedures.

The benefits afforded by any of these capabilities, however, must be balanced against those disadvantages generally associated with DBMS's and summarized below:

- High Initial Costs
  - Purchase of software, support, and training for the DBMS
  - Additional equipment costs - additional processing capability may be needed for the additional processing overhead, and storage requirements certainly will be increased over standard file maintenance (disk overhead factors of 100-300% are common for DBMS's depending on the structures of the data).
  - User training - few people are experienced with DBMS's and the complexity of these systems does not lend itself to self-instruction.
  - Conversion of existing data bases will cost both programmer and computer time.
- There is no standard DBMS (CODASYL's recommendations are very popular and are likely to provide the basis for any standard developed).
- Reliability is crucial because of the difficulty in providing good back-up procedures for a disk-resident data base. DBMS's have various schemes to allow "roll-back" to remove erroneous data entries but such schemes are quite complex and introduce much system overhead.

The high initial costs noted above probably constitute the most serious of the problems. It is rare that a new users' first application of a DBMS will save him money or effort because of these high costs. Once these costs are absorbed, however, operational costs may be very similar to those before implementating the DBMS. Computer operational costs will

probably rise since more processing and storage may be required, but programming costs may decrease.

Of course, the significance of any of the considerations noted above must be assessed in terms of their potential environment. DBMS's are still relatively young as commercial products. With time, their advantages will increase, no doubt with an accompanying reduction in disadvantages. Once a user decides the use of a DBMS appears appropriate, he should search for one with particular capabilities suited to his needs. The following subsection comments on DMS 1100 as a representative of well established DBMS's.

4.2 DMS 1100

DMS 1100 has been operational at various sites for a number of years. Without experienced assistance, however, we found the system very difficult to use (a characteristic we expect of most DBMS's because of their complex structures). Contracted support and training may be the most efficient approach to beginning use of this system. It is not anticipated that continued use of the system would be unreasonably difficult. Unfortunately, available documentation was much less than ideal and therefore new users would probably also require extensive training.

A definite advantage of DMS 1100 was its use of COBOL as a host language. With this configuration, the programmer is using a familiar medium and has all of its capabilities at hand along with the added features of the DBMS. As expected, processing overhead, however, was high and definite advantages must be demonstrated to counterbalance this cost.

PART 3

NASA DATA USERS' REQUIREMENTS FOR PROCESSING EQUIPMENT

Part 3 of this document categorizes potential user requirements for NASA data in terms of urgency, processing required, appropriate presentation media, and data retention requirements. Selected example users are then described in terms of these categories of requirements. This Part of the document further discusses these categories in terms of the effect of specific requirements on users' needs for processing equipment and suggests appropriate equipment to meet these needs.

1. INTRODUCTION

Prospective users of the scientific and geographic data originating from NASA programs will have a wide variety of data requirements, which will in turn necessitate a wide variety in the processing equipment appropriate for each user. Based upon an investigation of potential user data needs and the data acquisition and reporting plans for several NASA programs, this report discusses the associated user requirements for data processing equipment. The following Section addresses user data needs, while the final Section discusses the impact of these needs on equipment requirements.

2. USERS' DATA NEEDS

Because of the wide variety in the data needs of prospective NASA data users, it is impossible to describe a typical user and his data requirements. We can, however, examine the various types of data requirements expected of prospective NASA data users. Each potential user can be considered as a composite of various of these needs.

2.1 CATEGORIZATION OF DATA NEEDS

The following list represents a categorization of the requirements users will impose on the NASA data made available to them. It was selected to emphasize those aspects of using NASA data which would most influence how the data was made available to these users and what processing equipment these users would require.

- Urgency/Timeliness
- Processing Required
- Presentation of Data
- Data Retention Requirement

In the above context, urgency refers to how soon after a user identifies a need for certain data he must receive it. Timeliness refers to how old the data is before it is received. Thus, a user may have an urgent need for old historical data. In contrast, if a user has a need for timely data, his need is also urgent, since delivery time for the data effects its timeliness. Interractive real-time request and receipt of data is of course the most responsive answer to urgency and timeliness requirements. The next level of response would be that where users are

able to interractively identify and request appropriate data, but actual transmission of that data to the user is by correspondence. When data urgency and timeliness is of less concern, the mail service represents a convenient mode of communication.

The data processing that a user must provide in order to obtain the information he desires can range from none (an entirely manual system) to sophisticated data manipulation capabilities. Included within this range is the processing necessary to locate appropriate data (interractive directory assistance) and that which would allow a user to extract particular data elements from master data files.

The data presentation desired by a potential user will depend a great deal upon his need to consider spatial relationships. If this need is high, the need for graphical presentations will be high. Graphical presentations can also be used to effectively summarize data (e.g., bar graphs). Tabular presentations, however, are simpler to produce and virtually all users will desire some tabular output. Photographs represent still another data presentation available from certain NASA programs.

Data retention requirements refer to users needs to maintain historical data. Whereas some users will use only current data and have no need to retain it for future reference purposes, others doing projection and analysis studies will require extensive historical data files.

2.2 EXAMPLES OF POTENTIAL USERS

This subsection is included to illustrate how particular users can be classified by the above categorization of user needs. The example

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

users choosen are typical of those at the state and local levels of government who will be interested in NASA data. These specific examples illustrate that a single user's needs will vary from one application to another. Table 2-1 summarizes the data requirements of these users in terms of the categorization outlined above.

State and County agricultural agents represent a group of often cited potential LANDSAT data users. In particular, LANDSAT data can be used to determine much crop-related data, such as acreage of various crops, crop maturity, and levels of disease or insect infestation. To be of help for the harvest or protection of a current crop, this data must be made available as soon as possible. For analyzing acreage trends, however, historic data is necessary. In either case, large files of sensor data must be processed to identify those areas with the desired characteristics. This represents a heavy processing requirement and could mean hours of computing on a small system. Much of the data desired could be presented tabularly (i.e., number of acres of certain crops), and locations could be referenced by coordinates, but graphical presentations would also be useful. Photographs generated from the LANDSAT data are available and would be valuable in illustrating the geographic relationships of crops and terrain.

Land-use planners at the state, county, and local levels could .o use satellite data on land characteristics. Rather than the periodic needs of agricultural users, these planners would be interested in data on

USER	DATA REQUIREMENTS			
	URGENCY/ TIMELINESS	PROCESSING REQUIRED	PRESENTATION OF DATA	RETENTION OF DATA
Agricultural Agents	Critical for Crop Protection, Less Critical for Studies	Locate Appropriate Data (Much Central Processing to Extract Desired Data)	Tabular Graphical Photographic	Current for Monitoring Purposes, Historical for Studies
Land Use Planners	Data Must Present Current Land-Use	Locate Appropriate Data (Much Central Processing to Extract Desired Data)	Graphical Photographic	Primarily Concerned With Current Data
Conservationists	Critical for Monitoring Dangerous Conditions, Less Critical for Studies	Locate Appropriate Data (Much Central Processing to Extract Desired Data), and Local Manipulation for Studies	Tabular Graphical Photographic	Current for Monitoring Purposes, Historical for Studies

TABLE 2-1 Data Requirements For Selected NASA Data Users

particular areas as changes in land use for these areas are considered. Generally, such planning activities can tolerate lead time to acquire data, and the data itself need only be timely enough to reflect the current land use. Graphical displays would be desirable since the spatial relationships of phenomenon are in question. Generally, once the graphical representations of a particular area of interest are developed for specific characteristics, further processing is not required.

Conservation offlcals also can gain useful data from NASA programs. In particular, terain, vegetatian, and pollution levels can be observed from satelites. As a particular example, we can consider a conservationist studying beach erosion and river delta formation. Here, a combination of both LANDSAT and SEASAT data appears appropriate. Data on ocean currents, river flow, and shorelines could be examined for historical trends. Such a study may even call for projections into the future of these characteristics, possibly requiring added processing capabilities. Urgency of data acquisition would generally not be crucial for this user, but graphical presentations would no-doubt be necessary.

3. INFLUENCE OF DATA NEEDS ON EQUIPMENT REQUIREMENTS

In this Section, each of the categories of user reporting needs presented in Subsection 2.1 is discussed in terms of its impact on user equipment requirements.

3.1 URGENCY OF DATA

Urgency requirements for data will vary from needing data as soon as possible to being able to tolerate written correspondence and manual processing of data requests. Of course, no processing is required in this last case. If urgency is such that interractive inquiry or even interractive data retrieval is necessary, then processing needs can be met by a variety of devices. This equipment could be as simple as a teletype-like terminal with only communications capabilities, or as complex as a full computer system. The common component, however, would be the capability to communicate with a central site over transmission lines (such as phone lines).

3.2 PROCESSING REQUIRED

The equipment required for a user's local processing of NASA data will depend upon that user's total processing requirements and the amount of processing performed by NASA at central sites. The least amount of processing would be that sufficient to help users locate data pertinent to their application. A basic teletype-like user terminal supported by an interractive central directory service could meet this need. The next level of processing would include the capability for an individual user to obtain actual data over communication lines. The basic

teletype-like terminal mentioned above would be adequate at the local level for this application, but added central site capabilities would be necessary. Providing access to such a large store of data would be a major task, and any tabulation, summarization, or computational service would add to this effort. Once the user is able to receive data, he can perform virtually any level of local processing, provided he is able to justify the needed equipment. Generally, added local processing capabilities will allow the user to examine large quantities of data and physical media (i.e., magnetic tape) may be more appropriate than communication lines for data transfer between NASA and the local user.

3.3 PRESENTATION OF DATA

Tabular presentations of data will be the easiest to produce, but unfortunately are poor for presenting the spatial relationships which can be derived from such systems as LANDSAT and SEASAT. Virtually any hardcopy or CRT device can be used to display tabular data. Graphical data, however, requires a graphical CRT, a plotter, or a printer/plotter. Although color CRT's are available and pen colors can be changed in plotters, at this time these devices are generally limited to single color presentations with shading. If color presentations are needed, (i.e., for complex graphs or to depict multiple characteristics on a single geographic presentation) the more expensive color CRT's or the electronic generation of color photographs become necessary. The generation of such color photos from digital data requires very sophisticated equipment and is often available only through central service bureaus or in very large companies. The photographs taken by various NASA programs represent a final presentation approach, but have little impact on equipment needs.

3.4 RETENTION REQUIREMENTS

User's requirements for either current or historic data can influence their needs for various types of storage media. Current data is usually needed for a brief period of time, after which it may or may not be retained for historical purposes. For such applications of current data, large on-line storage media such as magnetic disk systems would facilitate manipluation of the data. For long-term storage of historic data, inexpensive non-volatile media is necessary. Currently, magnetic tape is a good media for such purposes. Often, needs for historic data are unpredictable and therefore, it would be desirable for NASA central sites to serve as repositories of historical data. To use this data, users would have to be equipped to receive and read data in the form and on the media supplied by NASA.